

The Temporality of Musical Experience: Philosophical Models and Embodiment

MARIA KON [1]

The University of Sydney, Unit for the History and Philosophy of Science and The Centre for Time

ABSTRACT: Recent philosophical work on temporal experience offers generic models that are often assumed to apply to all sensory modalities. I show that the models serve as broad frameworks in which different aspects of cognitive science can be slotted and, thus, are beneficial to furthering research programs in embodied music cognition. Here I discuss a particular feature of temporal experience that plays a key role in such philosophical work: a distinction between the experience of succession and the mere succession of experiences. I question the presupposition that there is such an evident, clear distinction and suggest that, instead, how the distinction is drawn is context-dependent. After suggesting a way to modify the philosophical models of temporal experience to accommodate this context-dependency, I illustrate that these models can fruitfully incorporate features of research projects in embodied musical cognition. To do so I supplement a modified retentionalist model with aspects of recent work that links bodily movement with musical perception (Godøy, 2006; 2010a; Jensenius, Wanderley, Godøy, and Leman, 2010). The resulting model is shown to facilitate novel hypotheses, refine the notion of context-dependency and point towards means of extending the philosophical model and an existent research program.

Submitted 2014 February 22; accepted 2014 October 8.

KEYWORDS: *modeling, succession, temporal experience, embodied cognition*

INTRODUCTION

OUR experience of music necessarily involves a temporal element: it occurs over time. Because of this tight link with temporality, in this paper I examine the possibility of relating the empirical investigation of musical experience to recent philosophical work on temporal experience. Such philosophical work offers discussion of temporal features of experiences as well as different competing but generic models that are usually presupposed to apply to all sensory modalities. While I believe that these features and models are in need of systematic empirically-informed refinement generally, here I argue that the models serve as broad frameworks in which different aspects of cognitive science can be slotted and, thus, are beneficial to furthering research projects in embodied music cognition.

To support these claims, I focus on an auditory version of a feature central to the philosophical models: the experience of succession (EoS). It is assumed that EoSs are sharply demarcated from successions of experience (SoEs). In EoSs there is supposed to be something extra experienced in addition to a succession of events. However, this gives rise to the question: what properties distinguish an EoS from a SoE?

Ian Phillips suggests one approach: EoSs, unlike SoEs, involve a "following on from" relation (2010, p. 178). This suggestion alludes to the postulation that EoSs are characterized as spanning a short interval of time. But, this approach is problematic: it is not clear what the duration of this interval is, especially regarding the upper boundary. I argue that whether one has an EoS is context-dependent because there are counterexamples in which being an EoS and having a particular duration come apart. In turn, I claim it is context that distinguishes EoSs from SoEs.

After stating how philosophical models can include this context-dependency, I illustrate that such models can fruitfully incorporate aspects of research programs in embodied musical cognition. To do so I supplement a philosophical model with elements of recent work that link bodily movement with musical perception (Godøy, 2006; 2010a; Jensenius et al., 2010). The resulting model is shown to facilitate novel, testable hypotheses, e.g., it can account for counterexamples by proposing that they involve different sound-motion links. Two such counterexamples are presented below. Additionally, the resulting model effectively refines the above general notion of context-dependency. Finally, in questioning the model's applicability to one of our cases, I am led to an extension of the model. I thus conclude that philosophical models of temporal experience can serve as catalysts for refining and extending research projects in embodied music cognition.

THE EXPERIENCE OF SUCCESSION AND THE SUCCESSION OF EXPERIENCES

A feature of temporal experience that is prominent in philosophical discussion regarding the modeling of such experience is a distinction drawn between the experience of succession (EoS) and the mere succession of experiences (SoE) (Kelly, 2005a, 2005b; Phillips, 2010; Dainton, 2014).[2]

In support of the intuition that there is such a distinction, this quotation from William James is often presented: "A succession of feelings, in and of itself, is not a feeling of succession. And since, to our successive feelings, a feeling of their own succession is added, that must be treated as an additional fact requiring its own special elucidation" (James, 1890, p.628). The contrast is then often exemplified by appeal to Broad's (1923) clock case. Consider seeing the hour hand of a clock move. To do so, you make a comparison between your present perception of the hand with your memory of the hand's position. From this you can then infer that it moved. Now consider seeing the second hand of a clock move. Compared to the hour hand, seeing this movement seems direct and non-inferential: you don't compare its positions and make an inference. The experience of hour hand movement is considered an SoE, while the direct, non-inferential experience of second hand movement is claimed to involve an EoS.

As for an auditory example, consider Phillips' (2010) main case: hearing a staccato broken C major chord played *allegro* on a piano. Call this "Case A". In contrast, suppose that the broken chord is played very slowly with, say, an hour interval between each note. Call this "Case B". Right now suppose you hear the G4 as played in Case B. At this time, you may recall the past notes. But, you cannot attend to all the notes occupying the interval. In this case, you do not have the EoS. Plus, I can make similar claims about these cases as are made about the clock case: your experience of the succession in Case A is direct and a comparison and inference is not required to attend to all the notes, unlike Case B. So, Case A, but not Case B, it is claimed, involves an EoS.

The quote and the clock example are then taken to support the assumption that there is some clear difference between the EoS and the mere SoE (Le Poidevin, 2007; 2009; Dainton, 2008; 2014; Kelly, 2005a, 2005b; Prosser, 2012). However, what exact features are relevant in distinguishing Case A from Case B? Or, more generally, what features distinguish an EoS from a SoE? Given the cases above, it seems that EoSs are characterized as being direct and non-inferential, while SoEs are not direct and require inference.

I find, however, that these features of being direct and non-inferential are problematically vague. To make salient the unsatisfactory nature of these features drawn from Broad's clock case, i.e., the features of "directness" and being apparently non-inferential, consider the following scenario. While at a jazz club you hear a lick by a solo tenor sax. Here you have the experience of a succession of notes. Contrast this experience with, while focused on the jazz band, hearing an assortment of thunks and clunks as patrons nearby return their glasses to their tables after taking a sip of their beverages at various times during the song. You probably don't consider these thunks and clunks to be a series of sounds in the same way that the lick is.

There seems to be some sort of difference in this case that may be described as a difference between the EoS and the mere SoE. However, I do not find it obvious how "directness" and non-inferentiality distinguish the way in which you experience the notes of the lick from the way in which you experience the clunking of glasses. Arguably, you had "direct" experience of the lick's notes as well as whatever clunking was occurring while the lick was played. Arguably, both could be considered non-inferential since, unlike in Case B and in perceiving an hour hand, one does not need to compare past experiences of previous sounds (or positions) in order to be aware of succession. In turn, I find that both of these features are too vague to be of use in drawing the distinction, e.g., it is not clear whether attention is a prerequisite to having "direct" experience, whether "directness" should incorporate an auditory figure-ground distinction, or what type(s) of memory is involved in the exercise of comparing experiences involved in making an inference.

Although I do not introduce the philosophical models of temporal experience until after elucidating this feature, a proponent of such a model may object that I am being too harsh: some of the models are developed such that this difference is indeed accounted for. For example, some (Crick and Koch, 2003; Le Poidevin, 2007) incorporate psychological mechanisms such that they give rise to the appearance of EoSs. Others (Dainton, 2000; Phillips, 2014) claim that experience is divided into temporally extended and overlapping units. It is this overlap that unifies the notes heard in Case A, but the notes experienced in Case B do not overlap and, thus, do not constitute an EoS.

In reply, while I do not deny that some of these models are geared to account for this difference, I want to step back and examine exactly the difference between EoSs and SoEs. So, rather than seeking to attack the models' solutions directly, I instead seek to refine the problematically vague distinction that these models attempt to account for. In turn, I shift the focus from how models accommodate this distinction to exactly what the distinction itself hinges upon. I argue here and in the

next section that the main features that philosophers suggest in this debate, regardless of whether they can be accommodated in existent models, do not straightforwardly ground the distinction.

Let's turn to answering this question: what features distinguish an EoS from a SoE?

THE CONTEXT-DEPENDENCE OF THE DISTINCTION

Some philosophers suggest other characteristics by which the distinction may be drawn. Although they, too, are vague, they do suggest that duration is the relevant demarcating feature. An EoS, as Dainton describes it, features a "temporal spread of contents" (2014, Section 1.2) presented together in consciousness such that they appear successive. Phillips (2010, p.178) characterizes an EoS as follows: you hear the G4 "follow on from" the previous two notes. Unlike Case B that features a set of momentary experiences that are separate and experientially isolated from each other, in Case A you can attend to this rapid series of notes that spans a short interval of time. So, in Case B, you just have the mere SoE. In Case A, there seems to be something extra experienced in addition to a succession of notes.

In turn, Phillips suggests that there is an additional relation—a following-on-from relation—experienced in EoSs that is not experienced in SoEs. So, perhaps your experience of the tenor's lick involves this following-on-from relation, while your experience of the clunking glasses does not. Although Phillips does not explicitly specify to what feature of the EoS this relation corresponds, the following examination of his cases point towards a suggestion that appears in both his and Dainton's descriptions, namely that it is a difference in duration that distinguishes an EoS from a SoE. However, after developing this duration feature, I argue that it, too, is insufficient to distinguish the experiences. Instead, I suggest that it is context that is the relevant feature.

Before discussing the duration feature, I must make a remark about my methodology and aim. In this section I aim to identify a general feature that grounds the EoS/SoE distinction. Although I do incorporate some more empirically-oriented data below, I try not to rely on specific theories, hypotheses or accounts at this stage. Instead, I wish to identify a *general* feature that can be incorporated in the philosophical models. In the next main section, I show how philosophical models can be modified such as to incorporate this general feature. It is not until the penultimate section below that I illustrate how these modified philosophical models may allow for the systematic integration of particular (and more heavily) empirically-informed accounts and models. In other words, in order to facilitate the application of philosophical models to more realistic cases of listening, i.e., those more akin to the jazz club case and less like the sparse and disembodied cases of hearing three piano notes, I must first ascertain a better *general* feature grounding the EoS/SoE distinction, a distinction which partially underpins these models.

The Following-on-from Relation and Duration

Case A is Phillips' paradigm of an EoS. Recall that in Case A, the notes are played very quickly. Suppose that you hear this series of notes and just now you hear the G4. According to Phillips, you hear the G4 "follow on from" the previous two notes. Unlike the slow Case B, which features a set of momentary experiences that are separate and experientially isolated from each other, you can attend to this rapid series of notes that spans a short interval of time. Thus, Phillips puts the temporal structure of the EoS in terms of some following-on-from relation in which its duration is somehow relevant: it spans a short interval of time such that you can attend to it.

Let's proceed to unpack this following-on-from relation in view of Case A's features and Phillips' discussion. Case A seems to have the following temporal features:

Feature 1 (F1): It involves the EoS of a particular duration, i.e., about a second.

Feature 2 (F2): It has a specific tonal center, i.e., C major.

F1 is discussed by Phillips, and his discussion does seem rely on F2 implicitly. I examine F1 in more detail below. However, I don't discuss F2 because it is particular to certain "musical" experiences of auditory succession, rather than experiences of auditory succession generally. Given my goal in this section of delineating a *general* feature that underpins the EoS/SoE distinction, I cannot simply adopt this feature to ground *all* EoSs, as EoSs that do not have a specific tonal center would be automatically ruled out. However, note that below I suggest that the feature is context-dependent. In turn, this broad context-dependency can be refined such as to allow F2 to be a relevant feature in particular contexts, in which case it will be important to incorporate an approach to tonality in one's model.

Regarding F1, Phillips (2010) explicitly wants to rule out the possibility that experiences of cases like Case B are EoSs. Plus, as mentioned above, he implies that experiences of succession are characterized as spanning a short interval of time. Other than a difference in total duration, both Case A and Case B have the same properties: they involve the same three notes played in the same order and spaced by intervals of equal duration.

But, how long or how short must an experience be to count as an EoS? The threshold below which we experience two successive auditory tones as simultaneous is 2-3ms (Wittmann, 2011, p. 2). So, this should be the lower boundary for experiencing two tones as successive. Yet, it's unclear what the upper boundary is.

Can some mechanism be appealed to in order to demarcate experiences of different durations? For example, can echoic memory be used as a means of demarcating experiences of different durations? Echoic memory lasts up to 2-5 seconds. It is the auditory version of sensory memory. Sensory memory is the shortest-term element of memory. It allows one to retain sensory information after the stimuli has ended (Bryne, 2003; Snyder, 2000). Suppose that Case C is played such that it doesn't quite fit in one's echoic memory: the echoic memory of C4 and E4 fades shortly before G4 is heard. It would still be in one's short-term memory: one would recall the C4 and E4, although the content of this recollection may be phenomenally less vivid.

However, other than this slight difference in duration and vivacity, Case C seems to have all the same properties as Case A if played slightly slower. Thus, it's still unclear what distinguishes a slightly slower Case A from Case C such that one has the EoS of the notes in one case and not the other. So, although one can in principle draw an upper boundary by appeal to different sorts of memory, it doesn't offer a means of distinguishing these cases in terms of their experienced temporal properties such that one case involves an EoS while the other does not.

Other than differences of duration, which, as I argued above, do not offer a clear means of delineating those that involve an EoS, what is it about the temporal structure of different cases that allows you to say that you had the EoS in one case?

The Context-Dependence of the Distinction

I claim that whether one has an EoS (above the 2-3ms threshold) is context-dependent. To motivate this claim, I present two cases.

The first case involves a postmodern concert. Suppose that you go to the concert of some famous postmodern composer. Only three notes are played. In fact, it's our favorite staccato broken C major chord played on a piano. C4 is played. Then, 10 minutes later, E4 is played. Finally, 10 minutes later, G4 is played. In the context of being in a postmodern concert, I am inclined to say that you do indeed have an EoS. This experience is something more than just the mere SoEs: there is some extra following-on-from relation among the notes. This seems due to the fact that you are in a particular context, and, given this context, you have certain expectations due to, perhaps, your past postmodern concert experience. In effect, this is a version of Case C that is an EoS due to the context in which it is heard.

My second case involves a cat and a cockroach. Suppose you see a cockroach scurrying across your baby grand. Your morbidly obese cat Fluffy is in pursuit. Fluffy swats at the cockroach a few times as it races across the keys. It turns out that you heard Fluffy play Case A. However, given this context of "observing nature", you have a mere experience of a succession of a C4, E4 and G4. In this case, you register that you heard three notes, but there's no following-on-from relation among them. Thus, it seems that Case A can be experienced without an accompanying EoS in certain contexts.

In view of these cases, EoSs aren't necessarily experiences that take place over a particular short duration. Moreover, other than the context, there isn't any apparent difference between, e.g., Case A as an EoS and as a mere SoE: both, for example, have the properties of being a succession of three notes and having a C major tonal center. Thus, rather than differences in duration, I suggest that it is the differences in context that go towards delineating an experience of auditory succession from a mere succession of experiences. It is important to emphasize that although the notion of context that I have introduced here is quite broadly construed, I show how it can be developed and refined in the penultimate section below.

In sum, Phillips assumes that cases that occur over longer durations cannot be EoSs. However, as I argued above, length of experienced duration (above a certain threshold) is not a temporal feature that allows you straightforwardly to distinguish cases in which you have the EoS from those in which you have a mere SoE. It seems, instead, it is context that goes towards delineating an EoS from a SoE. In turn, an account of the EoS must incorporate context.

PHILOSOPHICAL MODELS OF TEMPORAL EXPERIENCE AND CONTEXT-DEPENDENCE

In the previous section, I aimed to identify the general feature that distinguishes EoSs from SoEs. In this section I propose a general fix for philosophical models of temporal experience such that they can incorporate this general feature—the broadly construed notion of context-dependence. This section, like the previous section, aims to construct coherent models with very general concepts, rather than

attempting to systematically integrate empirical findings. The next section, however, shows how these generalized models can be refined using empirical findings.

Philosophical Models of Temporal Experience

Although there are several proposed classifications of philosophical models of temporal experience, I use that of Dainton (2011; 2014), which is the most widely accepted (Kon and Miller, 2015). He divides these models into three categories: cinematic, retentionalism and extensionalism. However, the cinematic view is not presented here due to space restrictions. To allow us to make salient the differences between these models, I adopt the general schema of the model's components suggested by Phillips (2010) as depicted in Figure 1.

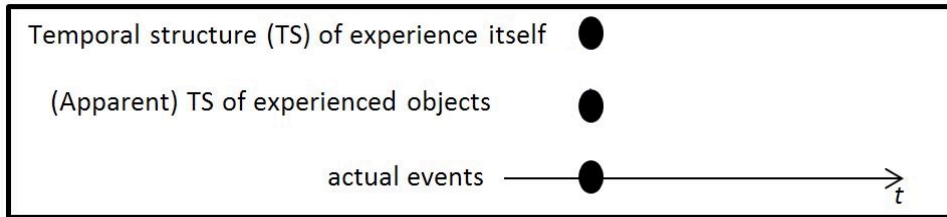


Figure 1. A general schema suggested by Phillips (2010) for representing the main philosophical models of temporal experience.

Suppose that a staccato quarter G4 is played on a piano. The bottom level labeled "actual events" provides the temporal structure of events that actually occur—just a note for a short duration at a particular time in this case. The middle "(apparent) temporal structure of experienced objects" level represents the temporal structure that you seem to experience, e.g., you seem to hear a lone G4. The top "temporal structure of experience itself" level represents the temporal structure of the experience itself. As will soon be made evident, some models assume that this level simply mirrors the structure of the middle level, while other models propose that the temporal structure of an experience is actually different from the way one experiences it, and it is this top level that represents the structure that the experience actually has.[3]

Figure 2 depicts the retentionalist model.

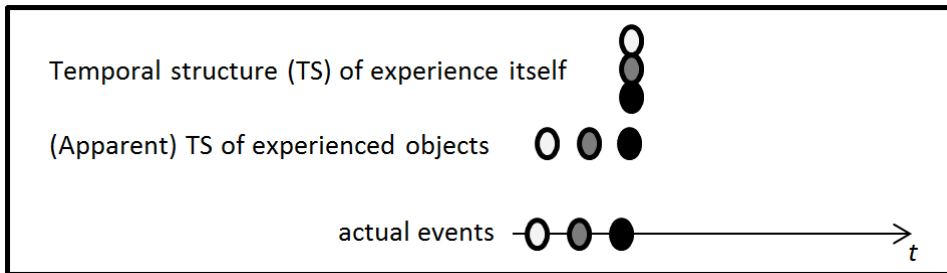


Figure 2. Dainton's (2014) Retentionalist Model.

This model attempts to give an account of the specious present—the doctrine that one has an apparently temporally extended experience at an instant of time (Le Poidevin, 2007; 2009; Dainton, 2014). According to retentionalism, a specious present is not actually extended over time. Instead, it involves some momentary states of consciousness that only *appear* to be spread over time. In terms of Case A, you have a mental representation of hearing C4, hearing E4 and hearing G4 simultaneously. This is depicted in Figure 2 at the top level. The top white oval at this level represents hearing a C4, the middle grey oval represents hearing an E4 and the bottom black oval represents hearing a G4. On the retentionalist model, the hearing of G4, in this case, is an immediate perceptual experience. The hearing of E4 and hearing C4 are retentions. Memories are supposed to be distinguished from retentions. Unlike memories, retentions are characterized as having "greater presence". This model can also be supplemented by Husserlian protentions; however, I do not include them here for the sake of simplicity.

The vertical depiction of the three ovals at the top level indicates that the mental representation of the three notes occurs at the same time. Since these mental representations are

simultaneous, it seems as though you should hear a chord. Yet, as depicted at the middle level in Figure 2, these tones appear to be experienced as successive.

One dominant mechanism that retentionalists use to account for this apparent temporal spread of these experiences is presentness: these experiences appear to be successive because the retentions have presence in different degrees. For example, your hearing C4 has less presentness than your hearing E4. One way of spelling this out is to claim that such degrees of presentness correspond to relative vivacity of retentions, e.g., your mental representation of C4 is less vivid than that of E4. There are problems with this use of presentness (Dainton, 2011; 2014), but I do not explicate or discuss them here.

Because I focus on this model below, I must make a few other notes about this traditional retentionism. Following Phillips (2010), I assume that such a model meets the following criteria. First, it must uphold the Principle of Simultaneous Awareness: if one experiences succession or temporal structure at all, then one experiences it at a moment. Second, the model must be compatible with the denial that irreducibly temporal facts—facts whose truth logically depends on states obtaining at times other than the present—are required in a satisfactory explanation of one's experience. For example, if one appeals to the fact that you had just heard C4, one is making recourse to your past state and, thus, to an irreducible temporal fact.

Contrast the retentionalist model with the extensionalist model depicted in Figure 3.

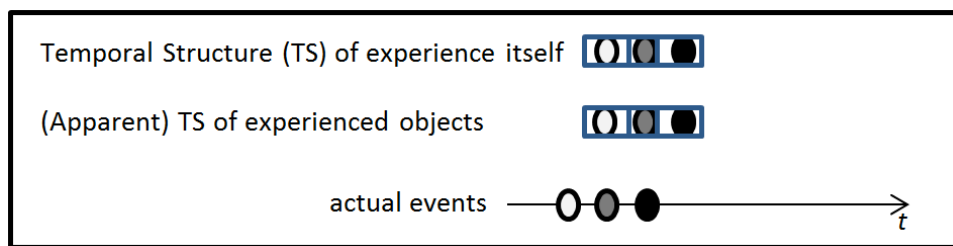


Figure 3. Dainton's (2014) Extensionalist Model.

This model has three main characteristics. First, it holds that experiential acts are temporally extended, rather than being instantaneous. It is assumed (Dainton, 2000, p.236) or established via introspection (Dainton, 2000, p.171) that these acts are of about 500ms in length. Notice the contrast between retentionalism: in retentionalism, experience is divided into phenomenal instants, with each experienced instant involving a perception and a set of retentions. However, an extensionalist regards experience as itself having temporal extension, which is divided into 500ms experiential acts. In the diagram, a single experiential act is represented by a pair of ovals, i.e., by a pair of heard notes, in a rectangle. Second, these extended acts overlap.[4] In Figure 3, the two overlapping rectangles represent two extended, overlapping acts. Third, unlike retentionalism in which the top and middle levels have different structure, the temporal structure of the experience itself just is that of the experienced objects.

As mentioned above and in Dainton (2014), these models propose some account of the EoS/SoE distinction. However, because I am more concerned with incorporating the feature that I have identified as central to the distinction, I neither present nor evaluate these accounts here.

The Projectionist Alternative

Although there are likely other general means of adapting the models given the context-dependent featured identified above, I propose the following Humean-style fix. To make the parallel salient, consider a typical Humean stance on causation. For the case of causation, suppose that there is no necessary connection between events. Rather, the idea of such a connection arises after seeing the constant conjunction of types of events. For example, one observes instances of the following two events on many occasions: putting bread in a toaster followed by toast popping out of it. In turn, one projects a causal relation between these events. However, this relation isn't really out in the world. Moreover, this relation is projected due to conditioning: one eventually projects this relation only after observing tokens of bread-into-toaster events followed by toast-popping-out events.

I formulate my alternative account of EoS along similar lines. You experience certain auditory events successively in certain contexts. For example, you hear C4s followed by E4s and G4s played on pianos at concerts. After such conditioning you come to perceive certain auditory experiences as EoSs. In such cases you project a succession relation among such experiences. For example, at a concert upon hearing a G4 preceded by an E4 and C4, you project a succession relation among them. It is the projection of such a relation that distinguishes the EoSs from the mere SoEs. In effect, given agent *A*'s

experience with content x in context C , a succession relation(s) is associated with x due to A 's conditioning.

To see how this fix applies to a model, consider retentionalism as outlined above. I determine whether the fix can be made compatible with the model by ascertaining whether retentionalism can be developed such that its criteria are met.

Does this use of conditioning use irreducibly temporal facts? Not necessarily: past states do not need to be appealed to in order to provide this account. Rather, one can appeal to one's present brain state and its wiring. One way of spelling this out is by appealing to a theory of learning from cognitive science. For example, consider Hebbian theory. The general slogan for this theory is that the neurons that fire together wire together; this theory is built on the claim that the connection between a pair of neurons is strengthened whenever these neurons are both active. In turn, through conditioning, one can develop and reinforce such neural wiring (Trappenberg, 2010). So, given certain stimuli, e.g., notes played at a concert at a time, and one's brain state and associated wiring at a time, one can make use of conditioning with reference to the present state only as follows. The neurons that are correlated with being at a concert are active at the same time as neurons that are correlated with hearing a series of tones in a certain way. Through conditioning, the connections among these sets of neurons are strengthened. In turn, given a seasoned concertgoer in the context of a concert, one can offer an explanation of the manner in which he or she experiences the succession of tones by recourse only to his brain state at a particular time. Thus, if supplemented with a theory about the learning process, the projectionist account can be developed such that it doesn't make use of irreducibly temporal facts.

Furthermore, if presented as a form of retentionalism, this alternative account can be depicted as in Figure 4.

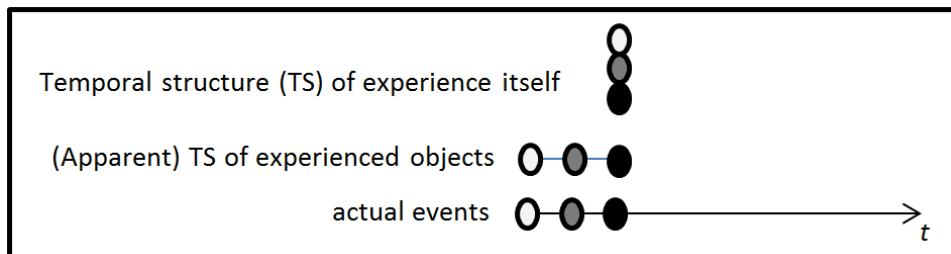


Figure 4. Projectionist Retentionalist Account of the EoS.

The temporal structure of experience itself is at a single moment. Thus, the Principle of Simultaneous Awareness is upheld. Further, to obtain the horizontal spread of events at the apparent temporal structure of experienced objects level, the retentionalist's degrees of presentness can be adopted. What distinguishes the projectionist retentional model from the traditional retentional model is the projected succession relation among experienced objects, which is represented by the lines connecting experienced objects at the middle level. Unlike the traditional retentional model, the projectionist retentional model has a means of incorporating the context-dependent EoS/SoE distinction: EoSs, unlike SoEs, are characterized by one's projection of a succession relation.

In sum, the alternative account upholds the Principle of Simultaneous Awareness. And, by appealing to one's brain state at a *particular* time, this account does not involve irreducibly temporal facts. Moreover, it offers an explanation of the context-dependant differences between EoSs and SoEs in terms of a projected succession relation: EoSs feature this projected relation, while SoEs do not. Whether this relation is projected is a function of one's brain state and one's set of experiences, retentions and/or memories one has at a particular time. Further, as with the causal relation in the Humean account of causation depicted above, the succession relation isn't actually out in the world and is projected due to conditioning. As such, the succession relation does not import irreducibly temporal facts. Thus, the projectionist fix can be applied to a philosophical model of temporal experience and, in turn, provide a general model that is sensitive to the feature of context-dependence that underpins the EoS/SoE distinction.

EMBODYING A PHILOSOPHICAL MODEL

To illustrate the ways in which the modified philosophical models of temporal experience can be of use in embodied music cognition, as well as to show how the broad notion of context-dependence may be refined, I supplement the projectionist retentionalist model with recent work by Godøy. Although I do not aim here to fit all of his account into this model, I do sketch how some of its key features may be incorporated such as to generate explanations of the postmodern concert and cat examples.

Here I adopt Godøy's definition of embodied cognition: "there is an incessant mental simulation going on in our minds of whatever we perceive, so that perception is not a matter of abstract processing of sensory data, but rather a process of reenactment of whatever we perceive" (2006, p.155).

Additionally, note that in places Godøy (2006; 2010a), and Godøy, Jensenius and Nymoen (2010) do advocate Husserlian views. As one version of the retentionalist model is that of Husserl (Dainton, 2014), this model may already be presupposed in parts of Godøy's work. Furthermore, Godøy (2010b) explores relations between Husserl's now-points and musical sound and movement. However, unlike Godøy's more historical inquiry that focuses on specific features of Husserl's account, I aim to show here that the general projectionist retentionalist model offers a useful framework for slotting in components of Godøy's work as well as other empirical accounts and hypotheses. Moreover, while Godøy (2010b) does highlight some links that Husserl suggests between now-points and action, I develop the retentionalist model in more detail below such as to incorporate motion.

The three components of Godøy's work that I use to sketch accounts of the cat and concert examples are his notions of motor-mimetic cognition, performance environment and chunks.

Godøy (2003) claims that there is a motor-mimetic component in music cognition, i.e., any sound will be included in some mental image of a gestural trajectory. This implies that there is a mental simulation of sound-producing gestures occurring when one perceives and/or imagines music (Godøy, 2006). So, there is a tight link between sound and motion generally, and in music between musical objects and musical gestures, e.g., between hearing Case A being played on a piano and imagining or mimicking the hand movements. Furthermore, Jensenius, Wanderley, Godøy and Leman refine the definition of "musical gesture" to mean, "an action pattern that produces music, is encoded in music or is made in response to music" (2010, p. 19).

While I do not present the different sorts of music-related actions that Godøy (2010a) and Jensenius et al. (2010) classify, I do highlight the role that the environment plays. In discussing different aspects of musical gestures, Jensenius et al. (2010) highlight the relevance of the performance environment: it places a constraint on the spatial aspects of musical gestures. They discuss this environment in terms of three concepts: performance scene, performance position, and performance space. The scene is a physical space that both the perceiver and performer recognize as one in which a performance is carried out, e.g., a clearly defined part of the stage in a typical concert situation, or the corner of Main and Poplar St. where a group of street musicians are performing. The position is the home position of a performer in which he or she sits or stands with the instrument prior to the performance. The perceiver usually knows, too, that the performance has not yet begun and does not expect sound to be produced while the performer is at this position. Finally, the space is an imaginary box surrounding a performer that defines the movement possible from a certain point in space. This box can be divided into different gesture spaces—imaginary boxes binding different musical gestures—e.g., a box around a piano's keys in which sound-producing hand movement takes place. In turn, the performance environment restricts musical gestures. Additionally, Jensenius et al. (2010) indicate that it induces particular expectations when perceiving a performance, e.g., one may experience surprise if a performer plays a piano by standing and plucking its strings.

Finally, Godøy (2011) claims that there are gestural-sonic objects—units that are based on the convergence of sound and music into holistically perceived chunks—that are approximately 3s long. He refers to these units as "chunks". Godøy (2011) gives several considerations as to why they are of this length, e.g., in music one may not be able to identify rhythmic, harmonic and melodic features of segments significantly less than 3s, while longer segments order chunks in a way that may be unimportant since a reordering of a piece's chunks may still be musical.

The components of this rough outline can be slotted into my projectionist retentionalist model. Godøy's chunks have a perceptual component, e.g., perceiving a sound, and each of these chunks has a motor component, i.e., mimicking or imagining actions associated with a sound. At a particular moment the perceptual component of a chunk can be identified with the relevant contents of one's immediate perception, retentions and short-term memory. Relevant features of the environment can also be incorporated into the retentional chunk. For example, consider the claim above that seeing the pianist in home position is a relevant feature. In turn, the retentional chunk must include visual elements of the performance space. To do so, visual perceptions are treated as I have done auditory perceptions: at a particular time there is an immediate visual experience of the pianist at home position, and some retentions of the seeing the pianist at home position.

Adding the motor component to these retentional chunks is not as straightforward. I suggest accommodating imagined movement and movement that one is aware of as follows: at a particular time add perceptions and imaginings of relevant bodily movement that has recently occurred to the perception, retentions and short-term memory. One's physical body movement is not as intuitively incorporable because the retentionalist model is geared towards accounting for perceptual phenomena and does so by breaking down one's experience into units that, although experienced as having some

short duration, are actually instantaneous sets of a perception and retentions. I suggest adding the position of one's body at a particular time to each of these sets. By stringing together these sets—at each instant there is a set comprised of one's body position, immediate perception, retentions and short-term memory—bodily motion is generated as one does with a flipbook. And, more importantly, this setup allows us at least in principle to break down bodily movement, whether physical or imagined, into a series of sets of bodily positions, perceptions, retentions and short-term memories. Thus, Godøy's chunks can be accommodated in the retentionalist model: the model associates physical body positions with mental contents at a particular time, and it also accounts for subjects' reports of their motion and experience of motion via the sets of perceptions, retentions and short-term memories. In turn, I have gone toward embodying a philosophical model: rather than being restricted to accounting for "the abstract processing of sensory data" (Godøy, 2006, p.155), this model also takes into account bodily movement.

Adding a Hebbian theory of learning, in addition to allowing us to meet the retentionalist criteria, may help develop how one comes to link particular motor, auditory and environmental features. Accordingly, after repeated exposure, one may come to associate types of motions, sounds and environments. This association results in specific expectations: you come to associate a particular stimulus or set of stimuli, e.g., a performance space of a concert featuring a solo pianist, sounds heard and imagined hand movements, with a set of notes that you group in a musical pattern. In turn, you come to expect to hear notes that will be grouped in a particular fashion given a particular environment and bodily motion. This, of course, also needs to be developed further and supplemented with an account of grouping, e.g., Snyder (2000). But, for my present purposes, this sketch suffices. Furthermore, if this speculative conglomeration of views can be supported and developed coherently, then it can afford a means of refining my broad notion of context-dependency: whether a succession relation is projected onto a series of, e.g., notes, is a function of the environment, sounds perceived and motion invoked.

With this sketch of a Godøy-informed retentionalist model, I can apply it to the cat and postmodern concert examples. First, consider the cat case in which your cat happens to play Case A in the process of pursuing an insect. Why might you not experience the notes as an EoS, while you would if Case A were performed by your neighbor? In view of the model, I hypothesize that you have different imagined/mimicked motions associated with the cases. In the cat case, you may imagine a general swatting motion, while you associate a fluid, human hand motion with the neighbor's playing. In turn, the latter results in a projected succession relation among the notes—they are grouped as a musical pattern—and, thus, constitute an EoS. In contrast, the notes heard in the former cat case may not be an EoS because you do not generally associate a swatting-insect motion with musical patterns and, in turn, you may not project such a relation among the notes.

My Godøy-informed model also offers an account of the postmodern concert example. Why are you inclined to connect the three notes, even though they occur 10 minutes apart, such as to have an EoS? I posit that it is due to expectations from being in a performance environment and the motion you associate with three notes. Here you form particular expectations due to being in a typical concert setting with a piano on a stage, and the performer sitting in a standard rest position before playing. Given this environment, you expect to hear notes in some sort of musical pattern. Moreover, given that it is a postmodern concert and assuming that you have experience of such concerts, you likely have different expectations as well, e.g., lower expectations of being entertained (Huron, 2006). Regarding the motor-sound link, you may correlate a motion with each of the notes, e.g., imagine pressing the key. Due to the expectations that arise given the environment plus, perhaps, your mimicking each note, you may link the notes together in some musical pattern and, thus, have an EoS, rather than merely regarding each note as a completely separate event.

However, one may worry that in this case it is the environment doing all the work: the motion-sound link hardly features here due to the lengthy silent intervals. In response, I suggest that there may be a sort of motion-sound association that features throughout the performance. During musical silences, i.e., periods of rest within a musical piece, one may be mimicking the position or motions, e.g., tapping foot, of the performer. Thus, one may be mimicking the poised position of the pianist and his or her glancing at a clock to some extent during the 10 minute intervals. Doing so may induce one to project a succession relation among the three notes.

From the above discussion, there are three upshots regarding prospects for my Godøy-informed model. First, the model accounts to some extent for the cat and postmodern concert examples. Moreover, in the process of applying the model to these examples, a novel, testable hypothesis was generated: I hypothesized that there are different motions associated with the cat and neighbor cases. Second, I refined my broad notion of context-dependence: it's a function of environment, motion and sounds perceived. Third, my discussion of the postmodern concert case has led me to formulate a possible extension of the model to musical silences.

CONCLUSION

I have shown how the generality of the philosophical models of temporal experience is advantageous in conducting empirical research because they can be supplemented and further developed by particular cognitive theories and findings. In effect, such models serve as general frameworks in which particular hypotheses and theories can be slotted and, thus, facilitate more unified and comprehensive accounts of embodied music cognition.

NOTES

[1] The author can be contacted at: m.kon19@gmail.com. She gratefully acknowledges funding from the John Templeton Foundation, via the project New Agendas for the Study of Time.

[2] I have identified the following three roles the distinction plays: (1) it is proposed that an adequate account of one's temporal experience must be able to account for the distinction (Kelly, 2005a, 2005b; Le Poidevin, 2007; Dainton, 2008; 2014); (2) it is offered as support for there being a specious present (Kelly, 2005a, 2005b; Dainton, 2014; Prosser, 2012); (3) it is used as an objection to a crude cinematic model (Kelly, 2005a, 200b; Le Poidevin, 2009; Dainton, 2008; 2014).

[3] See Dainton (2014) and Phillips (2014) for discussion of the relation between the middle and top levels.

[4] However, this overlapping characteristic is not found in all extensionalist views, e.g., Sprigge (1983). The overlap is included here because the recent, developed extensionalist accounts (Phillips, forthcoming; Dainton, 2000; 2014) feature overlap.

REFERENCES

- Broad, C. D. (1923). *Scientific thought*. London: Routledge.
- Byrne, J. H. (2003). *Learning and memory*. New York: Gale.
- Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6(2), 119-126.
- Dainton, B. (2000). *Stream of consciousness*. London: Routledge.
- Dainton, B. (2014). Temporal consciousness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2014 Edition). Retrieved from <http://plato.stanford.edu/archives/spr2014/entries/consciousness-temporal/>.
- Dainton, B. (2011). Time, passage, and immediate experience. In C. Callender (Ed.), *The Oxford handbook of philosophy of time* (pp. 382-419). Oxford: Oxford University Press.
- Dainton, B. (2008). The experience of time and change. *Philosophy Compass*, 3(4), 619-638.
- Godøy, R. I. (2003). Motor-mimetic music cognition. *Leonardo*, 36(4), 317-319.
- Godøy, R. I. (2006). Gestural-sonorous objects: Embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, 11(2), 149-157.
- Godøy, R. I., Jensenius, A. R., & Nymo, K. (2010). Chunking in music by coarticulation. *Acta Acustica*, 96(4), 690-700.
- Godøy, R. I. (2010a). Images of sonic objects. *Organised Sound*, 15(1), 54-62.
- Godøy, R. I. (2010b). Thinking now-points in music-related music. In R. Bader, C. Neuhaus, & U. Mergentstern (Eds.), *Concepts, experiments and fieldwork: Studies in systematic musicology and ethnomusicology* (pp. 245-260). Frankfurt: Peter Lang.
- Godøy, R. I. (2011). Coarticulated gestural-sonorous objects in music. In A. Gritten & E. King (Eds.), *New perspectives on music and gesture* (pp. 67-82). Surrey: Ashgate.

- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge: MIT Press.
- James, W. (1890). *The principles of psychology*. New York: Holt.
- Jenseni, A. R., Wanderley, M. M., Godøy, R. I., & Leman, M. (2010). Musical gestures: Concepts and methods in research. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 12-35). London: Routledge.
- Kelly, S. D. (2005a). The puzzle of temporal experience. In A. Brook (Ed.), *Cognition and the brain: The philosophy and neuroscience movement* (pp. 208-238). Cambridge: Cambridge University Press.
- Kelly, S. D. (2005b). Temporal awareness. In D. W. Smith & A. L. Thomasson (Eds.), *Phenomenology and philosophy of mind* (pp. 222-234). Oxford: Oxford University Press.
- Kon, M. & Miller, K. (2015). Temporal experience: Models, methodology and empirical evidence. *Topoi*, 34, 201-216.
- Le Poidevin, R. (2007). *The images of time: An essay on temporal representation*. Oxford: Oxford University Press.
- Le Poidevin, R. (2009). The experience and perception of time. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition). Retrieved from <http://plato.stanford.edu/archives/fall2011/entries/time-experience/>.
- Phillips, I. (2010). Perceiving temporal properties. *European Journal of Philosophy*, 18(2), 176-202.
- Phillips, I. (2014). The temporal structure of experience. In D. Lloyd & V. Arstila (Eds.), *Subjective time: The philosophy, psychology and neuroscience of temporality* (pp.139-158). Cambridge: MIT Press.
- Prosser, S. (2012). Why does time seem to pass? *Philosophy and Phenomenological Research*, 85(1), 92-116.
- Snyder, B. (2000). *Music and memory: An introduction*. Cambridge: MIT Press.
- Sprigge, T. L. S. (1983). *The vindication of absolute idealism*. Edinburgh: Edinburgh University Press.
- Trappenberg, T. (2010). *Fundamentals of computational neuroscience*. Oxford: Oxford University Press.
- Wittmann, M. (2011). Moments in time. *Frontiers in Integrative Neuroscience*, 5(66), 1-9.